

# **Self-Supervised Learning with Data Augmentations Provably Isolates Content from Style**

Year: 2021 | Citations: 348 | Authors: Julius von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel, B. Scholkopf

---

## **Abstract**

Self-supervised representation learning has shown remarkable success in a number of domains. A common practice is to perform data augmentation via hand-crafted transformations intended to leave the semantics of the data invariant. We seek to understand the empirical success of this approach from a theoretical perspective. We formulate the augmentation process as a latent variable model by postulating a partition of the latent representation into a content component, which is assumed invariant to augmentation, and a style component, which is allowed to change. Unlike prior work on disentanglement and independent component analysis, we allow for both nontrivial statistical and causal dependencies in the latent space. We study the identifiability of the latent representation based on pairs of views of the observations and prove sufficient conditions that allow us to identify the invariant content partition up to an invertible mapping in both generative and discriminative settings. We find numerical simulations with dependent latent variables are consistent with our theory. Lastly, we introduce Causal3DIdent, a dataset of high-dimensional, visually complex images with rich causal dependencies, which we use to study the effect of data augmentations performed in practice.